It is claimed:

1. A computer-implemented multi-dimension data analysis apparatus, comprising:

a computer data store for storing input data that has dimension variables

5   and at least one target variable;

a decision tree processing module connected to the data store that

determines a subset of the dimension variables for splitting the input data, wherein the

splitting by the dimension variable subset predicts the target variable; and

a multi-dimension viewer that generates a report using the determined

10   dimension variables subset and the splitting of the dimension variables.

2. The apparatus of claim 1 wherein the dimension variables subset includes continuous

variables.

15   3. The apparatus of claim 1 wherein the dimension variables subset includes category-

based variables.

4. The apparatus of claim 1 further comprising:

a selector module so that a user can alter which dimension variables to

20   include in the subset.

40

5. The apparatus of claim 4 wherein at least one statistic measure is provided to the user that is indicative of how well the splitting of the dimension variables predicts the target variable.

6. The apparatus of claim 5 wherein the statistic measure is a logworth statistic measure.

7. The apparatus of claim 1 further comprising:

a selector module so that a user can alter values at which the input data is split by the decision tree processing module.

8. The apparatus of claim 1 wherein the input data set includes a plurality of dimension variables and a single target variable.

9. The apparatus of claim 1 wherein the input data set includes a plurality of dimension variables and a plurality of target variables.

10. The apparatus of claim 1 wherein the decision tree processing module splits the input data into groups, wherein the multi-dimension viewer generates a report using the groups.

11. The apparatus of claim 10 wherein the decision tree processing module uses a competing initial splits approach to determine a subset of the dimension variables.

12. The apparatus of claim 11 wherein an initial split variable is indicated as most important variable in predicting the target variable.

13. The apparatus of claim 12 wherein a second split variable is indicated as second most important variable in predicting the target variable.

14. The apparatus of claim 1 wherein the decision tree processing module generates binary splits of the input data.

15. The apparatus of claim 1 wherein the decision tree processing module generates splits of the input data that are other than binary splits.

16. The apparatus of claim 1 wherein the generated report is viewed substantially adjacent to the dimension variables subset and the splitting values of the dimension variables subset.

17. The apparatus of claim 1 wherein the report has a format selected from the group consisting of a textual report format, tabular report format, graphical report format, and combinations thereof.

18. The apparatus of claim 17 wherein a marketing analyst selects one of the report formats in order to view the determined dimension variables subset and the splitting of the dimension variables.

42

19. The apparatus of claim 18 wherein the input data includes more than fifty dimension variables, wherein the determined dimension variables subset includes less than seven dimension variables that are viewed by the marketing analyst.

20. The apparatus of claim 1 wherein a user selects a type of summary statistics to view the determined dimension variables subset and the splitting of the dimension variables.

21. The apparatus of claim 1 further comprising:

a model repository for storing a model that contains the dimension variables and splitting values of the dimension variables.

22. The apparatus of claim 21 wherein the decision tree processing module splits the input data into a first set of groups according to first splitting rules to form a first model, wherein the decision tree processing module splits different input data into a second set of groups according to second splitting rules to form a second model, wherein the model repository includes a splitting rules index to store which splitting rules are used with which model.

23. The apparatus of claim 22 wherein the splitting rules index is searched in order to locate a model stored in the model repository.

24. The apparatus of claim 23 wherein the model repository includes a project level storage means, a diagram level storage means, and a model level storage means for storing the first and second models.

5   25. The apparatus of claim 22 wherein a search request is provided over a computer network to retrieve the first model from the model repository.

26. The apparatus of claim 25 wherein the computer network is an Internet network.

10   27. The apparatus of claim 22 wherein the model repository includes a plurality of specialty splitting rules indices that are used to locate a model stored in the model repository.

28. The apparatus of claim 27 wherein the specialty splitting rules indices are indices

15   selected from the group consisting of marketing specialty splitting rules indices, sales specialty splitting rules indices, and combinations thereof.

29. The apparatus of claim 22 wherein the model repository includes a mini-index means with a connection to the splitting rules index.

20

30. The apparatus of claim 1 wherein a data mining application provides construction of a process flow diagram, wherein the process flow diagram includes nodes representative of the input data and a variable configuration module.

31. The apparatus of claim 30 wherein an activated variable configuration module node provides a graphical user interface within which a user can alter which dimension variables to include in the subset.

5

32. The apparatus of claim 31 wherein the process flow diagram further includes a node representative of the decision tree processing module that has a competing initial splits approach for determining the subset of the dimension variables.

10   33. The apparatus of claim 31 wherein the process flow diagram further includes a node representative of the decision tree processing module that has a non-competing initial splits approach for determining the subset of the dimension variables.

34. A computer-implemented multi-dimension data analysis method, comprising the steps of:

storing input data that has dimension variables and at least one target variable;

5   determining a subset of the dimension variables for splitting the input data, wherein the splitting using the dimension variable subset predicts the target variable; and

generating a report using the determined dimension variables subset and the splitting of the dimension variables.

10   35. The method of claim 34 wherein the dimension variables subset includes continuous variables.

36. The method of claim 34 wherein the dimension variables subset includes category-based variables.

15

37. The method of claim 34 further comprising the step of:

altering which dimension variables to include in the subset of the dimension variables.

20   38. The method of claim 37 further comprising the step of:

providing at least one statistic measure that is indicative of how well the splitting of the dimension variables predicts the target variable.

39. The method of claim 38 wherein the statistic measure is a logworth statistic measure.

40. The method of claim 34 further comprising the step of:

    altering values at which the input data is split.

5

41. The method of claim 34 wherein the input data set includes a plurality of dimension variables and a single target variable.

42. The method of claim 34 wherein the input data set includes a plurality of dimension

10   variables and a plurality of target variables.

43. The method of claim 34 further comprising the step of:

    using a decision tree algorithm to determine the subset of the dimension variables by which to split the input data.

15

44. The method of claim 43 wherein the decision tree algorithm splits the input data into groups, wherein the multi-dimension viewer generates a report using the groups.

45. The method of claim 44 wherein the decision tree algorithm uses a competing initial

20   splits approach to determine the subset of the dimension variables.

46. The method of claim 45 wherein an initial split variable is indicated as most important variable in predicting the target variable.

47

47. The method of claim 46 wherein a second split variable is indicated as second most important variable in predicting the target variable.

5 48. The method of claim 34 further comprising the step of:

generating binary splits of the input data.

49. The method of claim 34 further comprising the step of:

generating splits of the input data that are other than binary splits.

10

50. The method of claim 34 wherein the generated report is viewed substantially proximate to the dimension variables subset and the splitting values of the dimension variables subset.

15 51. The method of claim 34 wherein the report has a format selected from the group consisting of a textual report format, tabular report format, graphical report format, and combinations thereof.

52. The method of claim 51 wherein a marketing analyst selects one of the report 20 formats in order to view the determined dimension variables subset and the splitting of the dimension variables.

53. The method of claim 52 wherein the input data includes more than fifty dimension variables, wherein the determined dimension variables subset includes less than seven dimension variables that are viewed by the marketing analyst.

5   54. The method of claim 34 wherein a user selects a type of summary statistics to view the determined dimension variables subset and the splitting of the dimension variables.

55. The method of claim 34 further comprising the step of:

     storing a model in a model repository, wherein the model contains the

10   dimension variables and splitting values of the dimension variables.

56. The method of claim 55 further comprising the step of:

     storing the model in a project level storage means, a diagram level storage means, and a model level storage means of the model repository.

15

57. The method of claim 55 wherein a search request is provided over a computer network to retrieve the model from the model repository.

58. The method of claim 57 wherein the computer network is an Internet network.

20

59. The method of claim 55 wherein the model repository includes a plurality of specialty splitting rules indices that are used to locate the model stored in the model repository.

60. The method of claim 59 wherein the specialty splitting rules indices are indices selected from the group consisting of marketing specialty splitting rules indices, sales specialty splitting rules indices, and combinations thereof.

5

61. The method of claim 34 wherein a data mining application provides construction of a process flow diagram, wherein the process flow diagram includes nodes representative of the input data and a variable configuration module.

10   62. The method of claim 61 further comprising the step of:

activating the variable configuration module node so that a user can alter which dimension variables to include in the subset.

63. A computer-implemented method for multi-dimension data analysis by a non-technical individual, comprising the steps of:

storing input data that has dimension and target variables;

receiving a request from the non-technical individual to analyze the stored

5 input data;

after receiving the request, determining a subset of the dimension

variables for splitting the input data, wherein the splitting using the dimension variable

subset predicts the target variable;

displaying the determined dimension variables subset and the dimension

10 variables so that the non-technical individual can alter which of the dimension variables

are included in the dimension variables subset; and

generating a report for the non-technical personnel using the dimension

variables subset as altered by the non-technical individual,

whereby the generated report is used for multi-dimension data analysis by

15 the non-technical individual.